The Report Committee for Jie Li
Certifies that this is the approved version of the following report:

Extension of the Cross-classified Multiple Membership Growth Curve Model for Longitudinal Data

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________

S.Natasha Beretvas

____

Peter Müller

Extension of the Cross-classified Multiple Membership Growth Curve Model for Longitudinal Data

by

Jie Li, B.S

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Acknowledgements

I would like to express my deep gratitude to Professor Beretvas, S.Natasha, my supervisor, for her patient guidance, enthusiastic encouragement and useful critiques of this master report. I would also like to thank Dr. Müller, Peter, for his advice and assistance.

Extension of the Cross-classified Multiple Membership Growth Curve Model for Longitudinal Data

Jie Li, M.S.Stat

The University of Texas at Austin, 2013

Supervisor: S.Natasha Beretvas

Student mobility is a common phenomenon in longitudinal data in educational research. The characteristics of education longitudinal data create a problem for the conventional multilevel model. Grady and Beretvas (2010) introduced a cross-classified multiple membership growth curve (CCMM-GCM) model to handle Student mobility over time by capturing complex higher level clustering structure in the data. There are some limitations in the CCMM-GCM model. By creating dummy coded indicators for each measurement occasion, the new model can improve the accuracy and provides an easier and more flexible structure at the higher level. This study provides some support that the new model better fits a dataset than the CCMM-GCM model

Keyword: Multilevel Model, CCMM-GCM, Student mobility, Cross-Classified Random Effects

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

Educational researchers are interested in the effects of different levels and the effects from various external factors that may have an influence on the student outcomes. The traditional statistical methods and techniques encounter difficulties to solve the problems concerning to the educational settings and research questions. In educational research, it typically is caused by the fact that variables of interest can be measured at different levels of the educational hierarchy. In other fields, these dilemmas are also encountered such as in medical research, biostatistics, and psychology and so on.

Hierarchical model or Multilevel model solves the "unit of Analysis "problem which is proposed by Bryk and Raudenbush (1998). With Hierarchical model, the effects of different level factors can be separated out and get assessment from observations such as the effects of class characteristics and the effects of schools. Hierarchical model also handles repeated measures on students allowing assessment of growth trajectories. The conventional Hierarchical model is mostly used for handling on the purely clustered data. In real situation, the movement of students is a general phenomenon. For example, by the report from U.S. Government Accounting Office (1994), more than 40% students changed their schools at least once from their first grade. A lot of research has been conducted to study the student mobility. For example, Kerbow (1996) studies the student mobility with a special curriculum or athletic program requirements. Kerbow (1996), Rumberger (2003) study the student mobility caused by external factors such as school environment.

The student mobility gives a difficult problem for education researchers. Typically, educational researchers will use two Ad-hoc methods to study the effects of student mobility: the first method is to delete the effects caused by the movement. Researchers will delete the data involved in student mobility or use a data set with pure cluster structure. This simplifies the question but reduces the power of analysis because the results is obtained by omitting some information from the data, for example, McRoach *et al* (2006) uses this method to study the children's reading ability; the second method is to simplify the data structure by just considering the one observation for individual student. For example, the researcher will keep the data that involves student mobility but only data from one school will be used to represent the student's performance. Obviously, the *first school* method has a lot of limitations such as incomplete information.

To solve the problem of student mobility, some new methods have been used such as Multiple Membership Random Effect Model (MMREM) and Cross-classified Random Effect Model (CCERM). For example, Goldstein *et al* (2007) use MMREM to study the student achievements; Grady *et al* (2010) provide a model combining cross-classified random effect model, Multiple Membership Random Effect Model and Growth model to study the performance of student over time. Though this model gives a way to study the longitudinal data set with considering the student mobility, it counts the effects from future state for current perform with a small weight.

In this report, a model based on Hierarchical model and Cross-Classified Random Effect Model is used to evaluate the students' performance over time. In the model, the effects of school, student mobility, and decayed effects from previous school are considered. This model will contain all information from the data set so the accuracy will be improved.

2 BACKGROUND

2.1 Hierarchical Model

Hierarchical model or multilevel model is commonly used in education field. For heuristic purposes, a two- level multilevel model is used for introduction.

At level 1,

$$\mathrm{Y_{ij}} = \pi_{0ij} + \pi_{1ij}\mathrm{X}_{ij} + e_{ij}$$

Where

$$e_{ij} \sim N(0, \sigma^2)$$

Level 1 model represents the relationship between individuals. In here, $\mathrm{Y_{ij}}$ means the outcome of student i who attends school j (i=1,….,N and j=1,…., M). For example, a student’s math score. π_{0ij} and π_{1ij} represent the initial state and growth parameter for student i.

At level 2,

$$\pi_{0ij} = b_{00} + r_{0j}$$

$$\pi_{1ij} = b_{01} + r_{1j}$$

Where,

$$r_{0j} \sim N(0, \sigma_1^2), r_{1j} \sim N(0, \sigma_2^2), cov(r_{0j}, r_{1j}) = 0$$

Level 2 model represents the relationship between clusters. For example, the effects from different schools. b_{00} is the grand mean for initial state and b_{01} is the grand mean for growth parameter. The random effects r_{0j} and r_{1j} are independent and assumed normally distributed with mean 0 and variance σ_1^2 andσ_2^2, respectively. This simple two level model can be used to assess the student's performance with pure hierarchical structure.

2.2 Cross-classified Random Effects Models

In conventional multilevel model, the data structure is assumed as pure cluster structure. For a pure cluster structure, the member of lower level units must belong to only one unit at the higher level. For example, student A attends Elementary school 1 and Elementary school 1 is a sub-unit of middle school 1 so student A will only attend middle school 1. The structure of pure clustering simplifies the problem in educational research. Students are nested in Elementary schools and Elementary schools are nested in Middle Schools. The conventional Hierarchical model provides an easy framework to solve questions in education field.

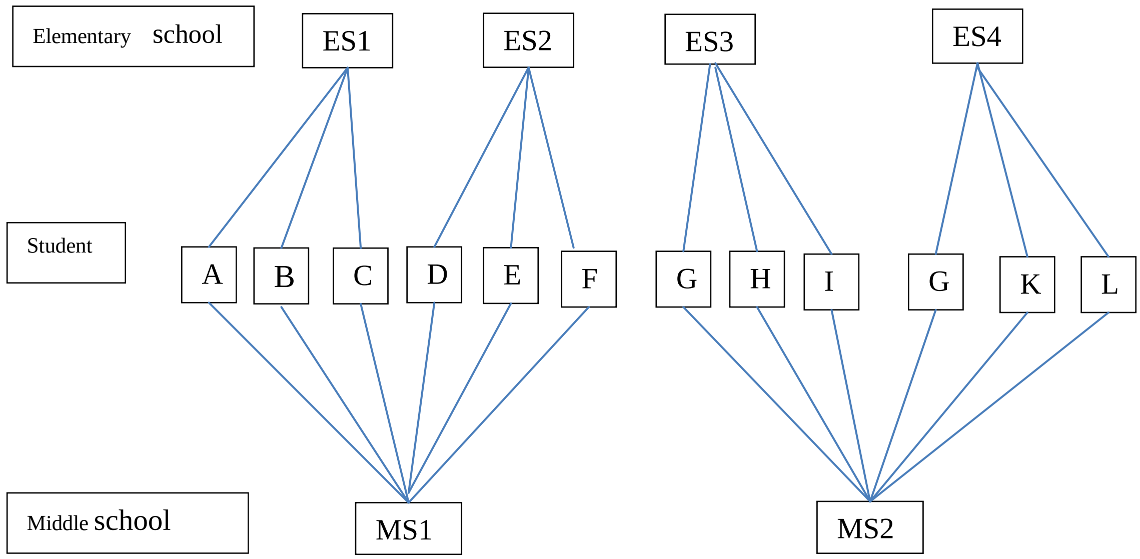

Figure 1 *networking graph to depict the pure hierarchical structure. The students are nested in elementary school and middle school.*

In real situation, the pure cluster structure is not realistic. Students may change school or move to other city. The movement of students is not a rare occurrence. The classical structure used for Hierarchical model is not appropriate for this situation. In the field of education, cross-classified structures are commonly encountered. The ad hoc methods used to prepare the data set for analysis will cause some problems. The *delete* method may mis-specify the model and possible give spurious results. The *first school* method can lead to an unnecessarily loss information and lack of statistical power. The Cross-Classified Random Effect Model provides a framework that can avoid the analytical issues and provides proper model structure.

For Cross-Classified structure, the member of lower level unit doesn't need to nest in on higher level unit (see Figure).

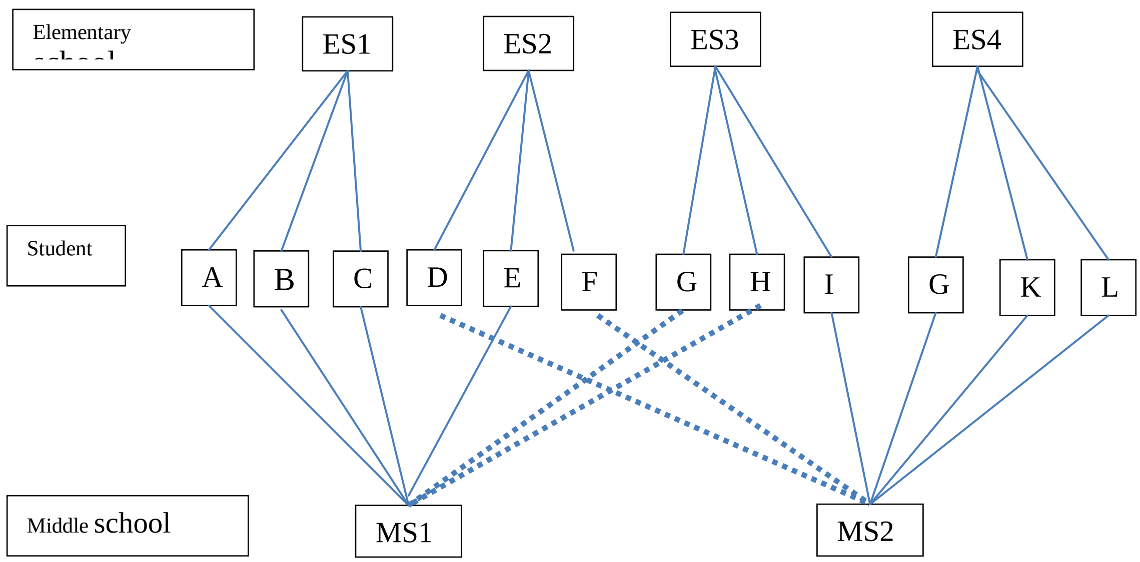

Figure 2 *networking graph to depict the cross-classified structure. The students are cross-classified by elementary school and middle school.*

An example derived from Raudenbush and Bryk(2002) is used to show the Cross-Classified structure. The example involves Elementary school and Middle school. Students (level 1) will attend elementary school and middle school, at level 1 the model is

$$Y_{i(j1,j2)} = \alpha + \pi_{0(j1,j2)} + e_{i(j1,j2)}$$

In the equation, $Y_{i(j1,j2)}$ is an outcome (such as performance) for student i who attends elementary school j1 and middle school j2. α is the grand mean for all students and $e_{i(j1,j2)}$ is the error term which follows normal distribution with mean 0 and variance σ^2, that is, $e_{i(j1,j2)}$~N(0, σ^2). The variable $\pi_{0(j1,j2)}$ represents the effects of combination of Elementary school j1 and Middle school j2. At level 2,

$$\pi_{0(j1,j2)} = \beta_{000} + \beta_{0j10} + \beta_{0j20} + e_{0(j1,j2)}$$

The β_{000} represents the grand mean for the combination effects, and $e_{0(j1,j2)}$ is the error term. β_{0j10} is the random effect from elementary school j1 and β_{0j20} is the random effect from elementary school j2. In the two –level CCREM, three subscripts are needed. The first subscript represents the level-one unit and subscripts in brackets represent the index of schools the student ever attended. In the second level, the effects from two schools can be separated into the effect from elementary school $\beta_{0j10}$and middle school β_{0j20} by partitioning the variability into two components.

2.3 CCMM-GCM

Educational researchers often want to assess the growth in the ability of students overtime. As mentioned in previous section, mobility of students is commonly encountered, especially for longitudinal data set. This creates a problem for conventional growth curve model (GCM). Grady and Beretvas (2010) create a framework to solve the problem by incorporating Multiple Member Random Effect Model (MMREM) and Cross-classified Random Effect model(CCREM) approaches into three level GCM to allow the intercept and slop of Growth Curve model to vary across time. They use a cross-classified structure to connect the effect of first school and subsequent schools. Within the subsequent schools, students are treated as member for each school in the subsequent set. A multiple membership structure is used to account the effect from subsequent school.

For CCMM-GCM model, the measurement occasion (level 1) is nested in students (level 2), students are nested in the schools (level 3) by a cross-classified structure. The students who attend more than one school will be a member of the subsequent schools, which is the second cross-classified factor. The model can be expressed as following:

At level 1

$$\mathrm{Y}_{it(\mathrm{j1},\{\mathrm{j2}\})} = \pi_{0i(\mathrm{j1},\{\mathrm{j2}\})} + \pi_{1i(\mathrm{j1},\{\mathrm{j2}\})} * \alpha_{1it(\mathrm{j1},\{\mathrm{j2}\})} + e_{it(\mathrm{j1},\{\mathrm{j2}\})}$$

At level2

$$\pi_{0i(\mathrm{j1},\{\mathrm{j2}\})} = \beta_{00(\mathrm{j1},\{\mathrm{j2}\})} + +r_{0i(\mathrm{j1},\{\mathrm{j2}\})}$$

$$\pi_{1i(\mathrm{j1},\{\mathrm{j2}\})} = \beta_{01(j1,j2)} + r_{1i(\mathrm{j1},\{\mathrm{j2}\})}$$

At level 3

$$\beta_{00(\mathrm{j1},\{\mathrm{j2}\})} = \delta_{100} + u_{0j1}$$

$$\beta_{01(\mathrm{j1},\{\mathrm{j2}\})} = \delta_{200} + u_{1j1} + \sum_{j\epsilon\{\mathrm{j2}\}} w_{tji} u_{1j}$$

The index $(\mathrm{j1},\{\mathrm{j2}\})$represents a cross-classified structure and index $\{\mathrm{j2}\}$ represents a multiple member structure for subsequent school set. $\mathrm{a}_{1ti(\mathrm{j1},\{\mathrm{j2}\})}$ is the time factor and it will be zero for the first time point or initial state. $\mathrm{w}_{tih}$is the weight for each school in subsequent school set. As all other Growth Curve model, CCMM-GCM model can include explanatory variables at each level such as gender in student level (level 2).

According to the result of Grady *et al*, the CCMM-GCM model provides a better result than first school approach with comparison of DIC. However, one limitation of CCMM-GCM is that it will count the effects from future schools into the current slope. For example, if researchers want to measure the growth of ability of students over a period containing three measurement occasions. At the second occasion, the slope will contain the effect from the 3^{rd} occasion which will happen in the future.

3 METHOD

3.1 Model specification

A dummied time factor Growth Curve model is provided to separate the student's effects over time. The structure of model will be simplified and multiple membership structure is not mandatory in the model. The slope in the model will be discrete while the slope in CCMM-GCM is assumed constant across time. The model combines hierarchical structure and cross-classified structure. The model can be expressed as following:

At level 1,

$$\mathrm{Y}_{\mathrm{tiJ}} = \pi_{1ij1} + \pi_{2i(j1,j2)} * I_2 + \pi_{3i(j2,j3)} * I_3 + e_{tij}$$

Where,

$$e_{ijt} \sim N\left(0, \sigma^2\right)$$

$\mathrm{Y}_{\mathrm{tiJ}}$ is outcome (math score) for student i who attends school J (J= j1,j2, j3) at the measurement occasion t . $\mathrm{j1, j2, j3}$ are, respectively, the index of school which the student attends at time t=1,2,3. Three time points are in the model and represent 2nd grade, 3rd grade, and 5th grade. π_{1ij1} is the initial state of student i. $\pi_{2i(j1,j2)}, \pi_{3i(j2,j3)}$ are the effects of school that the student i attends at time 2 and3.

At level 2

$$\pi_{1ij1} = \beta_{01j1} + + r_{1ij1}$$

$$\pi_{2i(j1,j2)} = \beta_{02(j1,j2)} + r_{2i(j1,j2)}$$

$\pi_{3i(j2,j3)} = \beta_{03(j2,j3)} + r_{3i(j2,j3)}$

Where,

$r_{0j} \sim N(0,\sigma_0^2), r_{1j} \sim N(0,\sigma_1^2), r_{2(j1,j2)} \sim N(0,\sigma_2^2), r_{3(j2,j3)} \sim N(0,\sigma_3^2)$

β_{000} is the grand mean for initial state and random effect can be added if needed such as the effect from kindergarten. $\beta_{01j}, \beta_{02(j1,j2)}, \beta_{03(j2,j3)}$ are the grand mean for effects from schools which the student attend at time 1,2 and 3. r_{kij}(k=0,1,2,3) are the random effects and assumed to be independent with each other.

At level 3,

$\beta_{01j1} = \delta_{100} + u_{1j1}$

$\beta_{02(j1,j2)} = \delta_{200} + u_{2j1} + u_{2j2}$

$\beta_{03(j2,j3)} = \delta_{300} + u_{3j2} + u_{3j3}$

Where,

$$\begin{pmatrix} u_{1j1} \\ u_{2j1} \\ u_{2j2} \\ u_{3j2} \\ u_{3j3} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \quad \right)$$

$\beta_{01j1}$for the assumption that the initial ability of students is same. $\delta_{100}, \delta_{200}, \delta_{300}$ are the grand mean for β_{01j1}, $\beta_{02(j1,j2)}$, $\beta_{03(j2,j3)}$, respectively. $u_{1j1}$is the effect from school j1 at time point1 , u_{2j2} is the effect from school j2 at time point 2, u_{3j3} is the effect from school j3 at time point2. u_{2j1} is the decayed effects of school j1 at time point 2. u_{3j3} is the decayed effects of school j2 at time point 3.,u_{1j1} and $u_{2j1}$are correlated that is ,cov(u_{2j1} ,u_{2j2}) ≠0 , u_{2j2} and u_{3j2} are correlated, that is, cov(u_{3j2} ,u_{2j2}) ≠0.

3.2 Prior distribution

For all fixed parameters and random components in the model, non-informative priors are used. This configuration will allow the data to dominate the inferences.

For all fixed parameters (δ_{100}, δ_{200}, δ_{300})non-informative normal distribution are used. The normal distribution has mean 0 and small precision. In Winbugs, the models are parameterized in terms of precision.

For random components the uniform distribution was used as priors for the precisions for level 1 and level 2. The range specified for the uniform distribution was 0 to 100.

For the random components in level 3, all random effects for schools follow multivariate normal distribution with mean 0 and variance $\sum$. $\sum$ is a variance and co-variance matrix and follows Wishart distribution. In the code, a conventional non-informative Wishart distribution is assigned to $\sum$ as the non-informative prior.

4 DATA SOURCE

The data source is from The Early Childhood Longitudinal Study (ECLS-K) which focuses on children experiences beginning with kindergarten and following children through middle school. The ECLS-K data provides descriptive information on student's performance and movement. The database records information from 21,409 students and it contains the measurements for the same children from kindergarten through the 8th grade. The data is collected in the fall and the spring of kindergarten (1998-99), the fall and spring of 1st grade (1999-2000), the spring of 3rd grade (2002), the spring of 5th grade (2004), and the spring of 8th grade (2007).

In the report, the measurements for math ability for 1st grade, 3rd grade, and 5th grade were extracted out and the information that records the movement of student in the three periods was also extracted. Data for students with missing values for any variables has been deleted. After cleaning the original dataset, the data set contained information from 1085 students.

5 RESULTS

All models were estimated using Winbugs (developed by members of Biostatistics Unit in Cambridge,UK) which is a Bayesian computation software based on Gibbs sampling. Two chains were used for each model and each chain is run with 50,000 iterations with a burn-in 20,000. To compare the performance of different models, the Deviance Information Criteria (DIC) is used. The definition is given by

$$DIC = Dbar + pD = Dhat + 2\,pD$$

Dhat is a point estimate of the deviance; pD is the effective number of parameters; Dbar is the posterior mean of the deviance. The model with the smallest DIC is considered as better model for current data set.

The results for fixed parameters differed somewhat. However, the values for fixed parameter in two models were similar. The average intercept in CCMM-GCM and in the model were almost the same: the average intercept for CCMM-GCM was 67.98; the average intercept for the new model is 65.48. For the average slope, the total average slop for the new model is 39.61 and 23.48 and the average slope for CCMM-GCM was 31.98.

Table 1 ***parameters and SE Estimates for new model***

Parameter	Mean	(SE)
Fixed parameter		
Average intercept	65.4	(0.5)
Average slope(total)		
Average slope at time 2	39.61	(0.4506)
Average slope at time 3	23.48	(0.143)
Random effects		
Level 1 variance		
Measure	63.45	(1.528)
Level 2 variance		
For intercept between student	244.8	(0.002)
For slope between student(total)		
For slope at time 2 between student	76.38	(4.424)
For slope at time 3 between student	5.155	(2.822)
Level 3 variance		
For intercept	26.12	(3.32)
For slope		
For slope at time 2		
The effect from previous school at time 2	14.58	(2.5885)
The effect from current school at time 2	0.089	(0.007)
For slope at time 3		
The effect from previous school at time 3	0.049	(0.004)
The effect from current school at time 3	14.38	(3.724)
DIC	22685.3	

The estimates for the random parameters appear remarkably difference between each other. The variance for measure is 63.45 while the result from CCMM-GCM is 107.6. The total variances for level 2 are 256.8 (for CCMM-GCM) and 326.3, respectively. The difference of model structure between two models causes a different partition of variance and total variance for the two models is almost same, 433.49 and 445.00. The DIC value for the new model is 22685.3 and the DIC value for CCMM-GCM is 24454.2. Thus, the new model is a better model for the dataset.

Table 2 ***parameters and SE Estimates for CCMM-GCM***

Parameter	Mean	(SE)
Fixed parameter		
Average intercept	67.98	(0.9755)
Average slope(total)	31.98	(0.4729)
Random effects		
Level 1 variance		
Measure	107.6	(4.046)
Level 2 variance		
For intercept between student	252.5	(13.96)
For slope between student	4.303	(2.969)
Level 3 variance		
For intercept between 1^{st} school	47.75	(11.72)
For slope between 1^{st} school	2.102	(1.435)
For Subsequent school	19.23	(5.883)
DIC	24454.2	

6 DISCUSSION

The goal is to provide an extended model based on CCMM-GCM to improve the fitting for dataset. From the results, the DIC from the model is smaller than the CCMM-GCM with same dataset. However, there are some limitations for the experiments and models.

First limitation is the sample size, though some pre-cleaning has been done to increase the sample size in each high level unit (such as school), the sample size for each high level unit is at least 15, still relatively small. Maas and Hox (2004) did some test on the robustness of estimates. They conclude that the variance in high level will be underestimated when the number of member in each group is less than 30. Under the model they used, the variance of the second-level slope variance was underestimated 3.1% if the group size was 5.The fixed parameters is robust for the sample size for high level unit. If a more ideal dataset can be used to test the two models, the result will be more confidence.

Second limitation is the prior distribution for variance-covariance matrix. The conventional "non-informative" version of Wishart or inverse-Wishart prior makes the marginal distribution of correlations as uniform and large standard deviations are associated with large correlations. The conventional form of Wishart is not a real informative prior. The method to solve the problem is based on separate strategy. Barnaud, MacCulloch, and Meng (2000) model the $\sum$ with $\Omega\Delta\Omega$ where Δ is the diagonal matrix of standard deviation and Ω is the correlated matrix. Another method is proposed by O'Malley and Zaslavsky (2008). They restrict the Ω to be positive semi-definite matrix and model the Ω and Δ simultaneously.

7 SUMMARY

The model incorporates more information from the data than GCM and provides a more flexible structure for the level2 and level3 with the dummied time factors while the multiple membership structure is not mandatory. For the different structure of models, the results from CCMM-GCM and the results from the model cannot be compared directly but the difference of DIC for two models indicates the model is better than CCMM-GCM.

APPENDIX

Appendix 1 Wingbugs Code for CCMM-GCM

```
#first school
model{
#level 1
for (t in 1:T){
y1[t]~dnorm(mu1[t],percision.y)
mu1[t]<-phi1[t]
y2[t]~dnorm(mu2[t],percision.y)
mu2[t]<-phi1[t]+phi2[t]*1
y3[t]~dnorm(mu3[t],percision.y)
mu3[t]<-phi1[t]+phi2[t]*2
}
#level 2&level3
for (i in 1:T){
phi1[i]~dnorm(beta10[i],percision.phi1)
beta10[i]<-gamma100+u10[ sch[i,1],1]
phi2[i]~dnorm(beta11[i],percision.phi2)
beta11[i]<-gamma110+u10[ sch[i,1],2]+0.5*u12[sch[i,2]]+0.5*u12[sch[i,3]]
}
#parameters
mu[1]<-0
mu[2]<-0
for (j in 1:J){
u12[j]~dnorm(0,percision.u12)
u10[j,1:2]~dmnorm(mu[1:2],percision.u10[1:2,1:2])
}
s1[1,1]<-1
s1[1,2]<-0
s1[2,1]<-0
s1[2,2]<-1
percision.u10[1:2,1:2]~dwish(s1[,],10)
#percision.u1<-inverse(temp.percision.u1)
sigma.u10[1:2,1:2]<-inverse(percision.u10[1:2,1:2])
temp.percision.u12~dunif(0,100)
percision.u12<-pow(temp.percision.u12,-2)
sigma.u12<-1/percision.u12
gamma100~dnorm(0,1.0E-5)
gamma110~dunif(0,100)
temp.percision.phi1~dunif(0,100)
percision.phi1<-pow(temp.percision.phi1,-2)
```

```
percision.phi2~dgamma(0.01,0.01)
#percision.phi2<-pow(temp.percision.phi2,-2)
sigma.phi1<-1/percision.phi1
sigma.phi2<-1/percision.phi2
#level1
temp.percision.y~dunif(0,100)
percision.y<-pow(temp.percision.y,-2)
sigma.y<-1/percision.y
}
```

Appendix 2 Winbugs Code for the new model

```
model{
#level 1
for(t in 1:T){
y1[t]~dnorm(mu1[t],percision.y)
mu1[t]<-phi1[t]
y2[t]~dnorm(mu2[t],percision.y)
mu2[t]<-phi1[t]+phi2[t]
y3[t]~dnorm(mu3[t],percision.y)
mu3[t]<-phi1[t]+phi2[t]+phi3[t]
}
#level 2
for(i in 1:T){
phi1[i]~dnorm(beta1[sch[i,1]],percision.phi1)
phi2[i]~dnorm(beta2[sch[i,1],sch[i,2]],percision.phi2)
phi3[i]~dnorm(beta3[sch[i,2],sch[i,3]],percision.phi3)
}
#level 3
for(j1 in 1:J){
beta1[j1]<-gamma000+u1[j1,1]
for(j2 in 1:J){
beta2[j1,j2]<-gamma100+u1[j1,2]+u2[j2,1]
beta3[j1,j2]<-gamma200+u2[j1,2]+u3[j2]
}
}
mu[1]<-0
mu[2]<-0
#define parameters
for (j3 in 1:J){
u1[j3,1:2]~dmnorm(mu[1:2],percision.u1[1:2,1:2])
u2[j3,1:2]~dmnorm(mu[1:2],percision.u2[1:2,1:2])
u3[j3]~dnorm(0,percision.u3)
}
s1[1,1]<-1
s1[1,2]<-0
s1[2,1]<-0
s1[2,2]<-1
percision.u1[1:2,1:2]~dwish(s1[,],15)
#percision.u1<-inverse(temp.percision.u1)
s2[1,1]<-1
s2[1,2]<-0
```

```
s2[2,1]<-0
s2[2,2]<-1
percision.u2[1:2,1:2]~dwish(s2[,],15)
#percision.u2<-inverse(temp.percision.u2)
temp.percision.u3~dunif(0,100)
percision.u3<-pow(temp.percision.u3,-2)
sigma.u1[1:2,1:2]<-inverse(percision.u1[1:2,1:2])
sigma.u2[1:2,1:2]<-inverse(percision.u2[1:2,1:2])
sigma.u3<-1/percision.u3
gamma000~dnorm(0,1.0E-5)
gamma100~dnorm(0,1.0E-5)
gamma200~dnorm(0,1.0E-5)
#level 2
temp.percision.phi1~dunif(0,100)
percision.phi1<-pow(temp.percision.phi1,-2)
temp.percision.phi2~dunif(0,100)
percision.phi2<-pow(temp.percision.phi2,-2)
temp.percision.phi3~dunif(0,100)
percision.phi3<-pow(temp.percision.phi3,-2)
sigma.phi1<-1/percision.phi1
sigma.phi2<-1/percision.phi2
sigma.phi3<-1/percision.phi3
#level 1
temp.percision.y~dunif(0,100)
percision.y<-pow(temp.percision.y,-2)
sigma.y<-1/percision.y
}
```

REFERENCE

Barnard, J., McCulloch, R., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage.*Statistica Sinica*, *10*(4), 1281-1312.

Beretvas,S.N.(2010). Cross-classified and multiple membership models. In Hox, J. J., & Roberts, J. K. (2011). *Handbook of advanced multilevel analysis*. London:Routledge.

Beretvas,S.N.(2008). Cross-Classified random effects models. In O'Connell, A. A., & McCoach, D. B. (2008). *Multilevel modeling of educational data(pp.161-197).* Information Age Pub Incorporated.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, 65-108.

Choi, K., & Seltzer, M. (2010). Modeling heterogeneity in relationships between initial status and rates of change: Treating latent variable regression coefficients as random coefficients in a three-level hierarchical model. *Journal of Educational and Behavioral Statistics, 35*(1), 54-91.

Gelman, A. (1995). Method of moments using Monte Carlo simulation. *Journal of Computational and Graphical Statistics*, *4*(1), 36-54.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, *1*(3), 515-534.

Grady, M. W., & Beretvas, S. N. (2010). Incorporating student mobility in achievement growth modeling: A cross-classified multiple membership growth curve model. *Multivariate Behavioral Research*, 45(3), 393-419.

Goldstein, H. (2011). Multilevel statistical models (Vol. 922). Wiley.

Goldstein, H., Burgess, S., & McConnell, B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(4), 941-954.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. Journal of Education for Students Placed at Risk, 1(2), 147-169.

Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*(2), 127-137.

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. Journal of Educational Psychology, 98(1), 14-28.

O’Malley, A. J., & Zaslavsky, A. M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*, *103*(484), 1405-1418.

Rumberger, R. W. (2003). The causes and consequences of student mobility. Journal of Negro Education, 6-21.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*(3), 263-286.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

U. S. General Accounting Office. (1994). Elementary school children: Many change schools frequently, harming their education. (GAO/HEHS-94-45, pp. 1–55). Washington, DC:U.S. Government Printing Office.